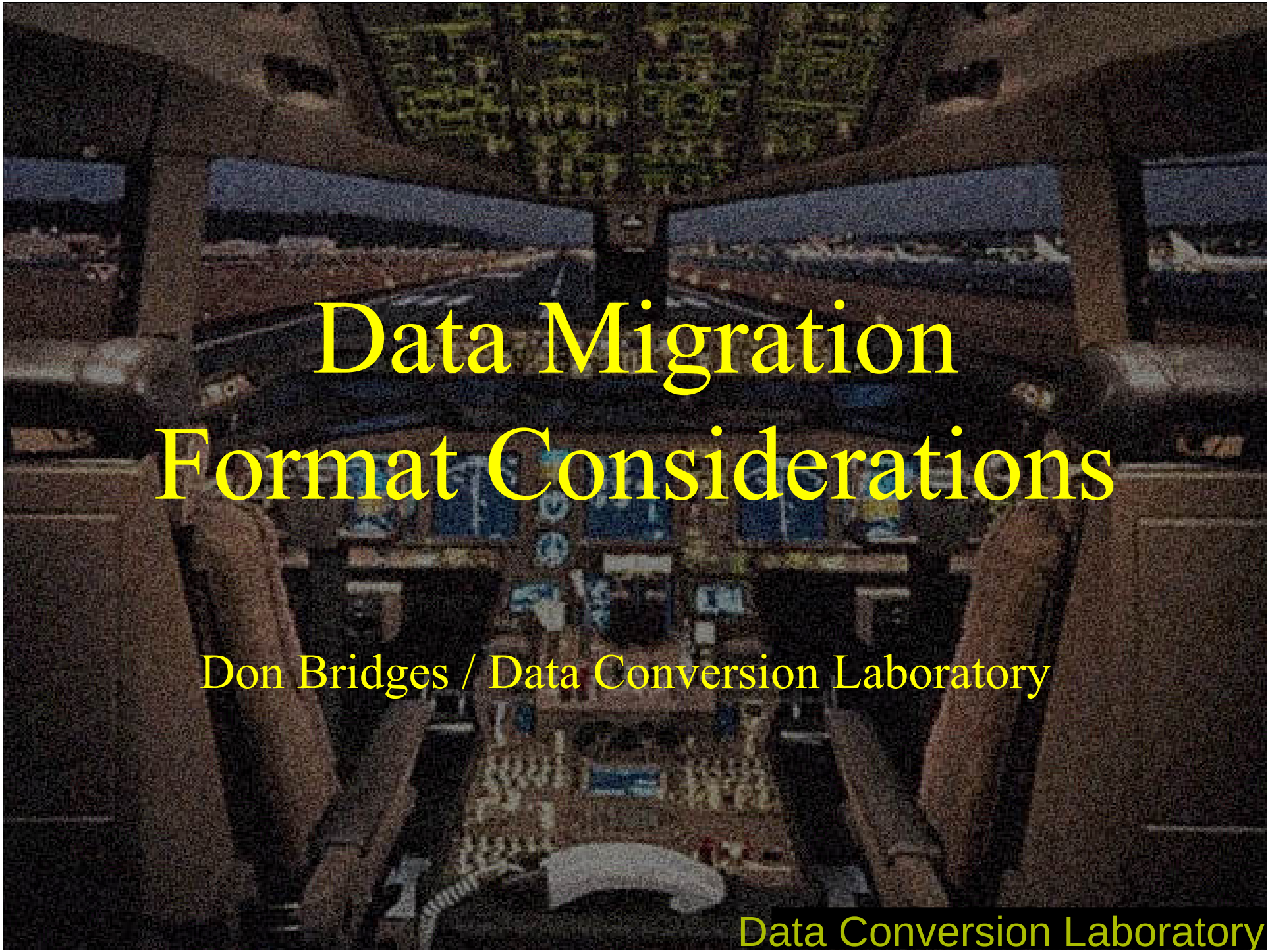
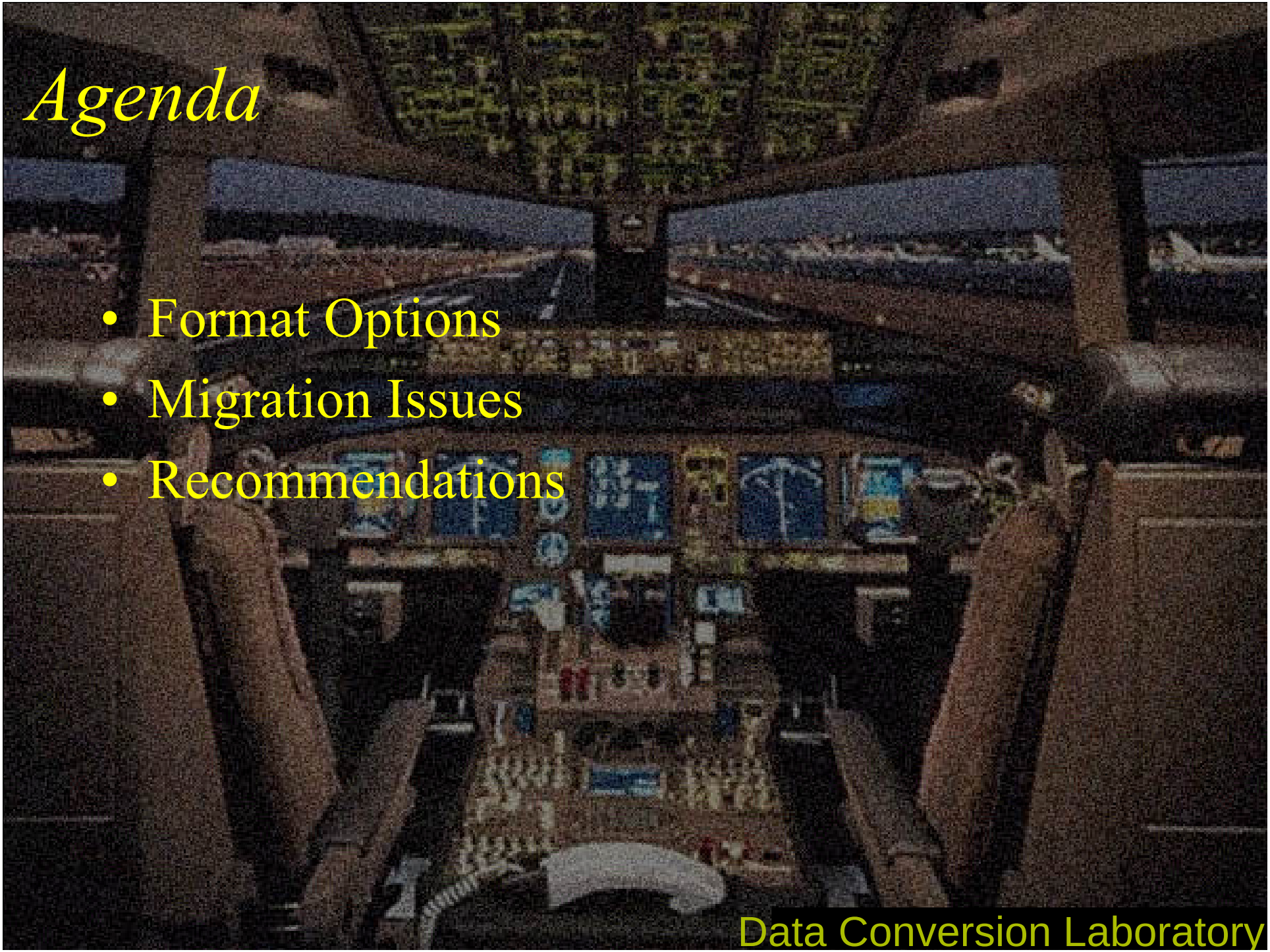


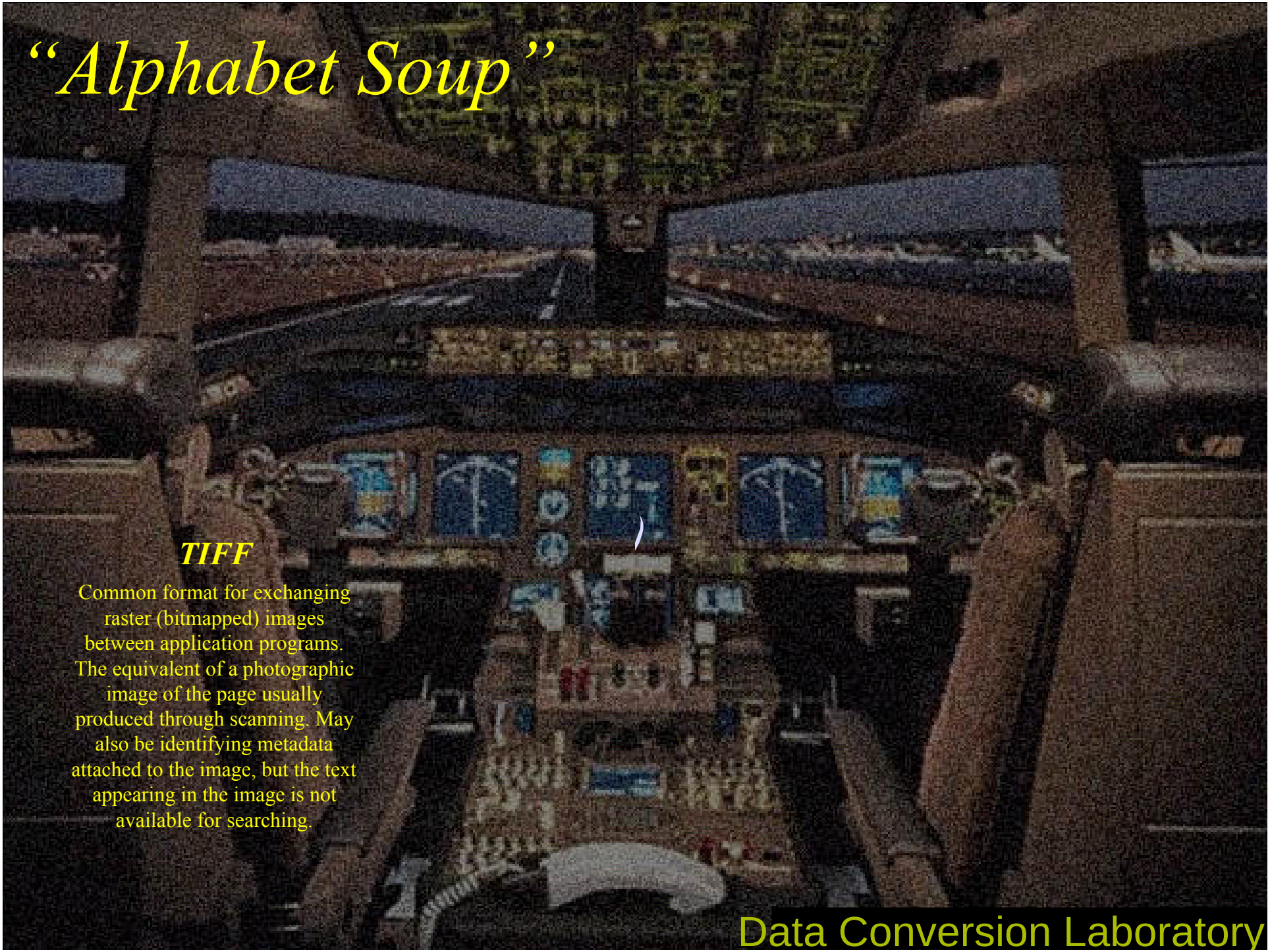
Data Migration Format Considerations

Don Bridges / Data Conversion Laboratory



Agenda

- Format Options
- Migration Issues
- Recommendations



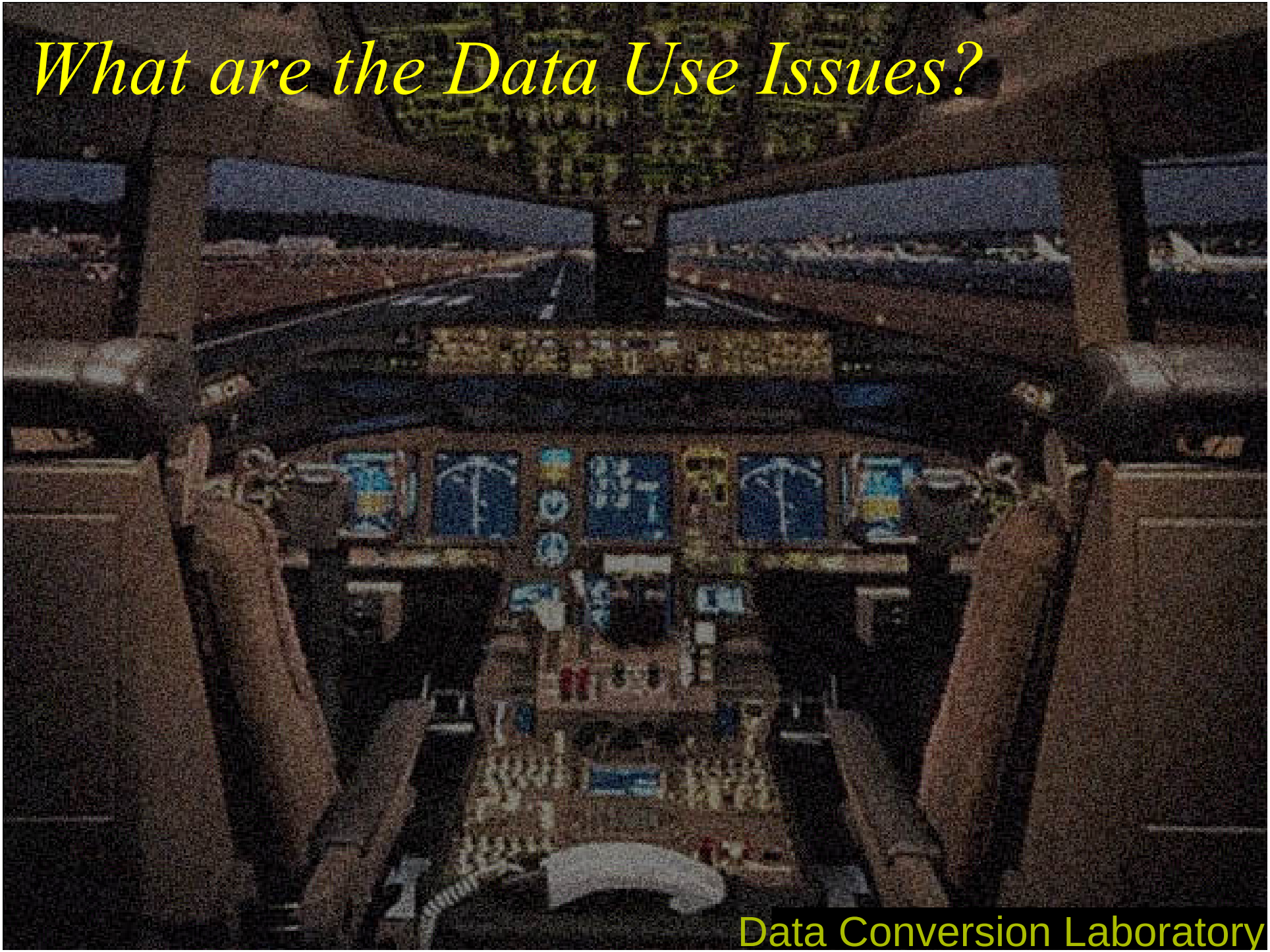
"Alphabet Soup"

TIFF

Common format for exchanging raster (bitmapped) images between application programs. The equivalent of a photographic image of the page usually produced through scanning. May also be identifying metadata attached to the image, but the text appearing in the image is not available for searching.

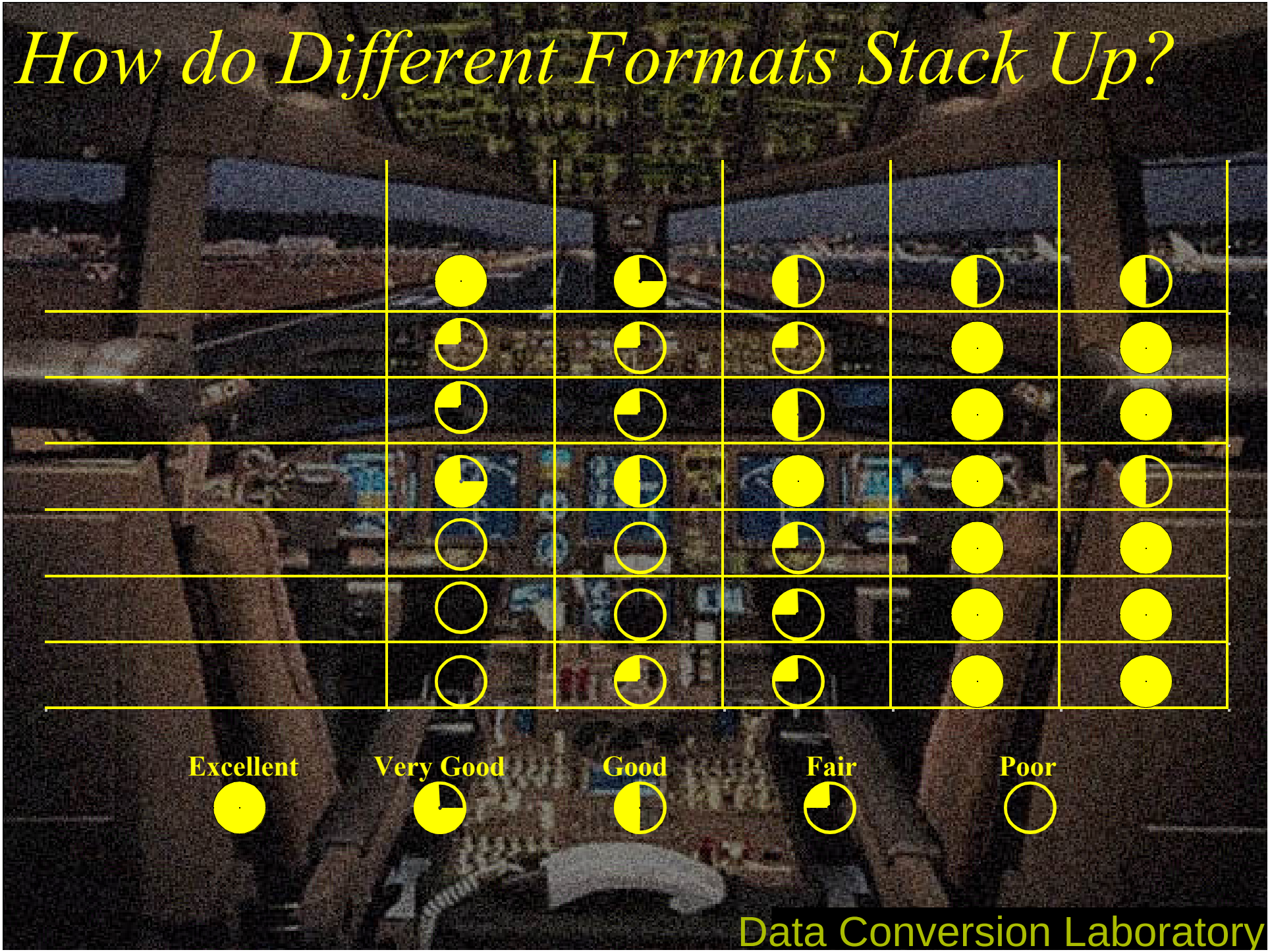
Data Conversion Laboratory

What are the Data Use Issues?



Data Conversion Laboratory

How do Different Formats Stack Up?



Data Conversion Laboratory

DTDs and Schemas

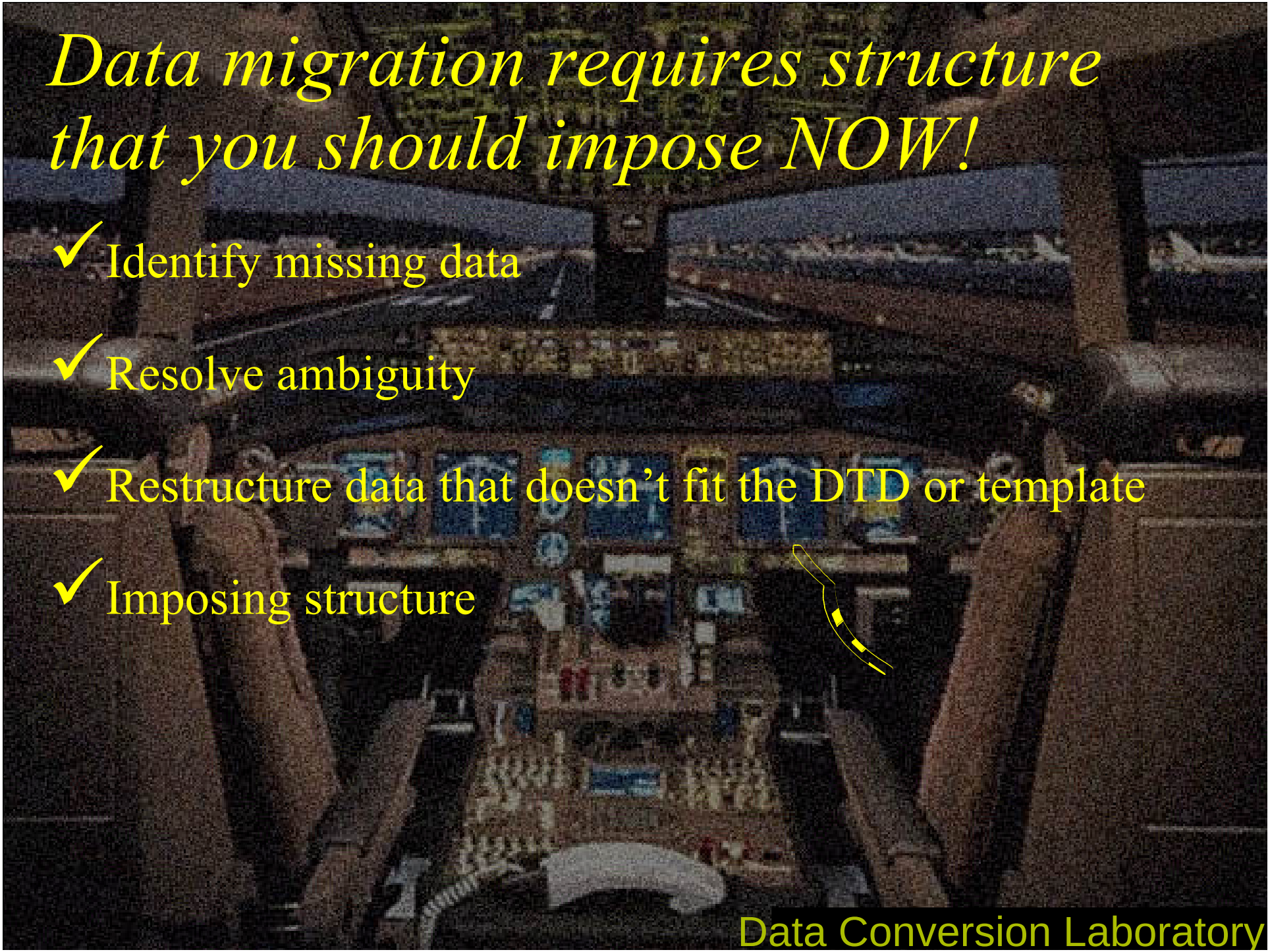
- DTDs define the structure of data
- Schemas add rules to structure definitions
  - primarily currently used for database applications
- Either way, your DTD or Schema should:
  - Account for your legacy data
  - Leverage off of industry standards
    - ATA TICC
  - Meet anticipated internal requirements

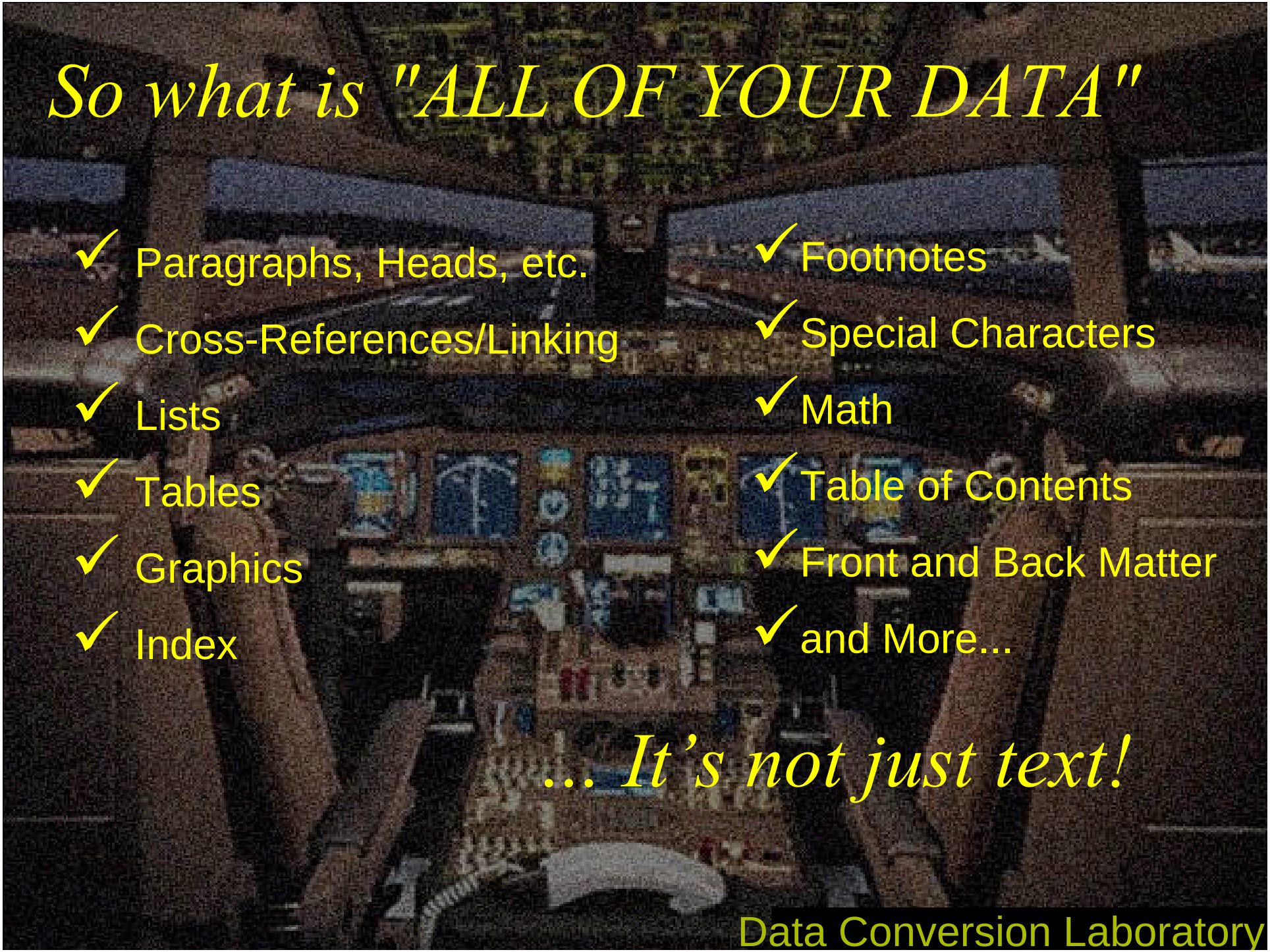
SGML or XML ?

- New to Mark-up languages? Go with XML
  - ~all the advantages of SGML
  - More and Cheaper tools
- Already established in SGML? Stay there
  - You've done the hard work
  - Your tools are are in-place
  - But – make you SGML "XML compliant"
    - SGML data should be XML or near-XML compliant resulting in "painless" conversion to XML

Typical conversion issues

- Quality
  - When you do show it to the world, will it make you proud?
- Time to Market
  - Can the world wait a year for you to be ready?
- Cost
  - Do you know what it really will cost? Are you sure?
- Scalability
  - Can you do thousands or millions of them the way you did your demo?


Data migration requires structure that you should impose NOW!
✓ Identify missing data
✓ Resolve ambiguity
✓ Restructure data that doesn't fit the DTD or template
✓ Imposing structure
Data Conversion Laboratory


So what is "ALL OF YOUR DATA"
✓ Paragraphs, Heads, etc.
✓ Cross-References/Linking
✓ Lists
✓ Tables
✓ Graphics
✓ Index
✓ Footnotes
✓ Special Characters
✓ Math
✓ Table of Contents
✓ Front and Back Matter
✓ and More...
... It's not just text!
Data Conversion Laboratory

The "why don't you just ..." trap

Example: Hyperlinks

- See figure 15.5
- See fig. 15.5
- Refer to figure below
- As illustrated on previous page
- Sentenced to 15.5 years…
- See figure 15.1 in volume II of …

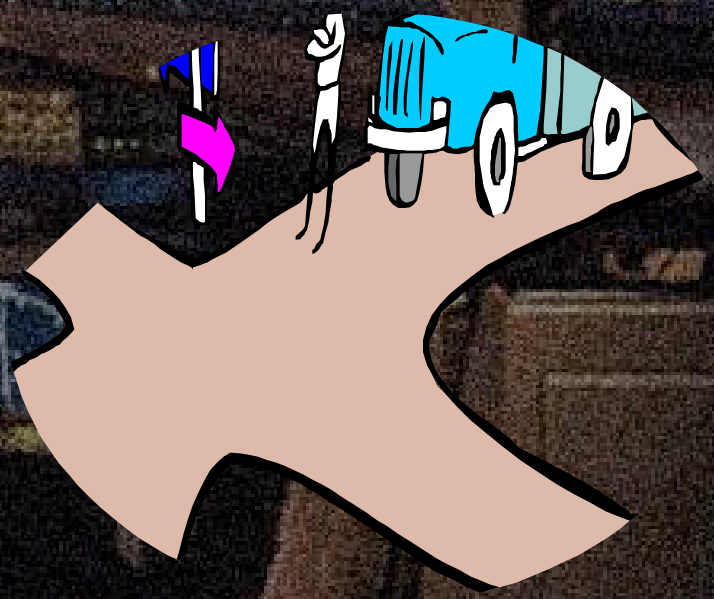
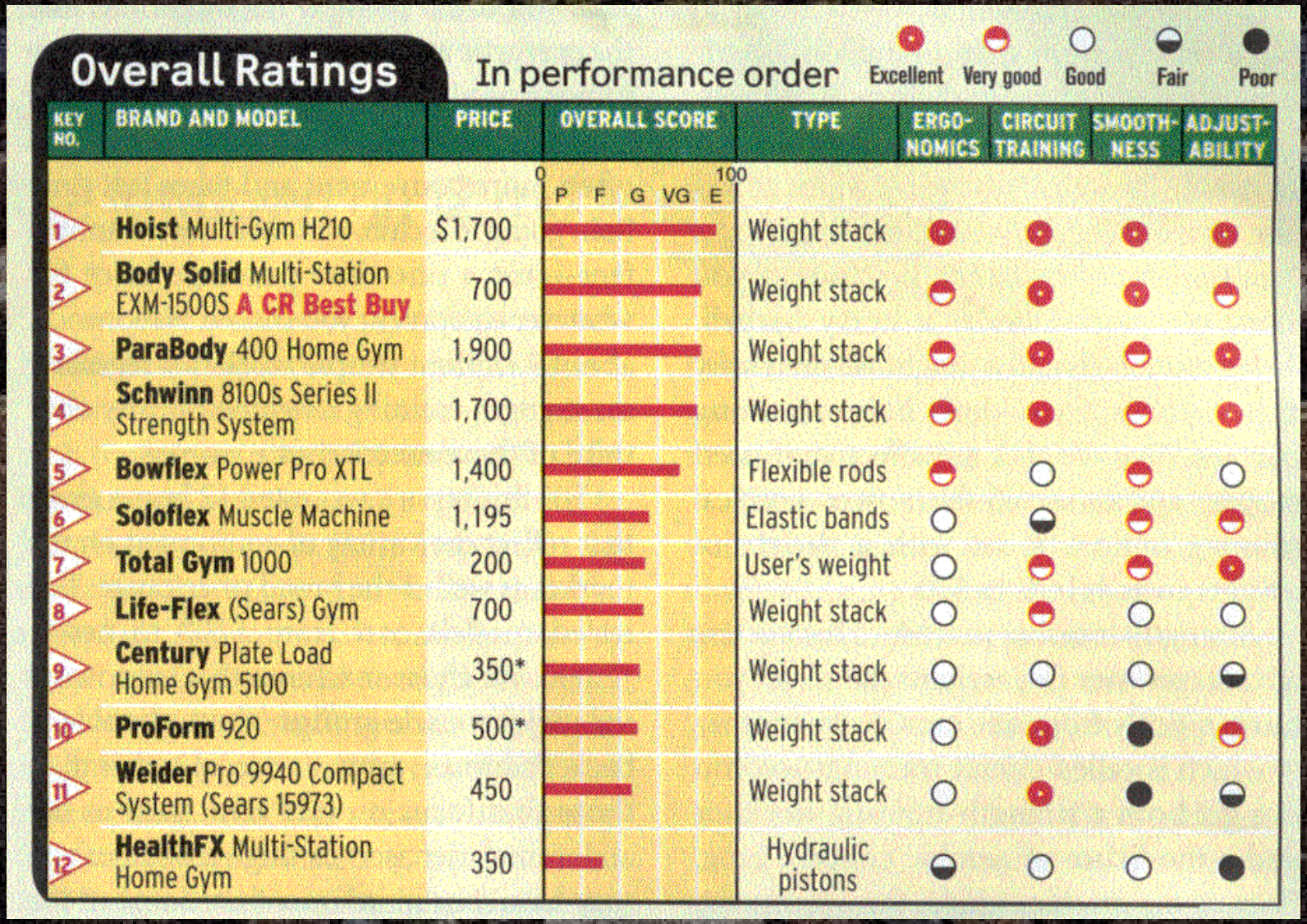


Table Example



Overall Ratings

In performance order

Excellent · Very good · Good · Fair · Poor

KEY NO.	BRAND AND MODEL	PRICE	OVERALL SCORE (0–100; P F G VG E)	TYPE	ERGONOMICS	CIRCUIT TRAINING	SMOOTHNESS	ADJUSTABILITY
1	**Hoist** Multi-Gym H210	$1,700		Weight stack	Excellent	Excellent	Excellent	Excellent
2	**Body Solid** Multi-Station EXM-1500S **A CR Best Buy**	700		Weight stack	Very good	Excellent	Excellent	Very good
3	**ParaBody** 400 Home Gym	1,900		Weight stack	Very good	Excellent	Very good	Excellent
4	**Schwinn** 810Os Series II Strength System	1,700		Weight stack	Very good	Excellent	Very good	Excellent
5	**Bowflex** Power Pro XTL	1,400		Flexible rods	Very good	Good	Very good	Good
6	**Soloflex** Muscle Machine	1,195		Elastic bands	Good	Fair	Very good	Very good
7	**Total Gym** 1000	200		User's weight	Good	Very good	Very good	Excellent
8	**Life-Flex** (Sears) Gym	700		Weight stack	Good	Very good	Good	Good
9	**Century** Plate Load Home Gym 5100	350*		Weight stack	Good	Good	Good	Fair
10	**ProForm** 920	500*		Weight stack	Good	Excellent	Poor	Very good
11	**Weider** Pro 9940 Compact System (Sears 15973)	450		Weight stack	Good	Excellent	Poor	Fair
12	**HealthFX** Multi-Station Home Gym	350		Hydraulic pistons	Fair	Good	Good	Poor

How low do you want to go …

Example: Granularity of Information

Printed Reference:

Actuator, Assembly of
P/N4076300
Implemented by SB 4076A

Example 1:

<part><partname> Flap, Nozzle Convergent </partname>
<partnumber> 4076300 </partname>
<sb> 4076A </sb>
</part>

How low do you want to go …

Example: Granularity of References

Printed Reference:

Actuator, Assembly of
P/N4076300
Implemented by SB 4076A

Example 2:

```
<part hazmat="no">
<partname> Flap, Nozzle Convergent </partname>
<partnumber> 4076200 </partname>
<manf> Pratt & Whitney (/manf>
<material> stainless steel </material>
<coating> acrylic enamel </coating>
<sb> 4076A </sb>
<effective> 89-776 </effective>
</part>
```

Not all lists are created equal …

Example: Numbered Lists vs. Series of Steps

- Do you need to differentiate between a simple numbered list and a series of steps? They appear similar to software, however, they have different meanings
- Structure is different
  - ✓ Steps typically have more rules … use the power of technology

Sequential List Example

Relubricate the following:

1. Flap tracks #1 and #2 in each wing.
2. Landing gear and doors.
3. Baggage door handle shaft.

Steps Example

If gust lock found engaged, proceed as follows:

1. Check control wheel and column for excessive movement.
2. Examine flight control surfaces for distortion.
3. Examine elevator stop bracket for

A picture is worth a thousand words

Example: Graphics Conversion

- Which format should you use?
  - ✓ Raster (TIFF, GIF, JPEG, PNG, etc.) – less expensive, static
  - ✓ Vector (CGM, SVG, IGES, AutoCad, etc.....) – more expensive, hot spotting, editable, re-sizing, less space
- Resolution?
- How good is your source?
  - ✓ electronic or paper?
  - ✓ is source vector?

Keep your heading above water …

Example: Document Headings

- Heading Hierarchy is not always obvious
- Heading Hierarchies are often complex
  - ✓(sometimes going 8 – 10 levels deep)
- Heading levels may not always be consistent within the same document
  - ✓be aware of skipped levels
- Converting multiple documents while trying to normalize headings
  - ✓different levels of granularity
  - ✓Mapping different heading styles

Which way should you go?

Use in-house conversions if...

- Your schedule is flexible
  - You'll need to expect the unexpected
- Materials are not complex
  - You'll be involved with post-conversion clean-up requirements that may vary widely
- Budget is tight & in-house resources available
  - You'll purchase tools that are relatively cheap
- Project is small
  - You'll need about one week of clean-up per ~500 pages (at 5 minutes a page), plus any project set-up efforts.

Which way should you go?

Use out-source conversions if...

- Meeting schedule is critical
  - You'll have a realistic estimate of how long the project will take as well as your options for speeding it up
- Materials are complex
  - Data conversion experts will have experience dealing with the difficult and/or unusual issues
- Budget is well defined
  - You'll have an understanding of the project costs and what the trade-offs are
- Project is large
  - You'll have a process that scales as big as you want

Recommendations

- Select format(s) that meets your expected requirements
  - If a functionality is not required, why pay for it?
  - Rookies should consider XML, veterans stay with SGML
- Define a DTD/Schema that looks backwards and forwards
  - Accounts for condition of legacy data
  - Leverages efforts of industry standards
  - Addresses internal requirements for future use & reuse
- Migrate data based on your situation
  - Your current conversion does not have to be the ‘final’ answer
  - Budget / Quality / Schedule constraints